Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE DETERMINATION OF SECULAR TREND[1]

By W. L. Crum, *Yale University*

It is customary in the analysis of ordered statistical series of the historical economic sort to eliminate the growth factor, assumed constant, by fitting a straight line to the original data. The equation of this straight line is determined by the method of least squares [2] or by an equivalent process. At a later stage in the investigation, it is frequently desirable to consider the ratio of the deviation of a particular variate from the ordinate of the corresponding point on the line of secular trend to that ordinate, rather than the actual deviation itself.[3] The question arises whether it would not be preferable to fit the line so that the sum of the squares of these *ratios*, rather than the sum of the squares of the actual *deviations*, is minimum. It is the purpose of this paper to examine this question and to get a first approximation to the position of the line of secular trend on the new basis.

I. THE COEFFICIENT OF CORRELATION

Suppose we have a series of N—N being assumed odd for convenience—variates x_i ordered relative to a variable t_i. We suppose further that the intervals between successive t_i are all equal and equal to unity, and we therefore represent the t_i as i; and moreover we take the middle t_i as origin. Finally, we restrict the problem to cases in which the secular trend is substantially rectilinear and gradual, and in which the total change in x is not a large fraction of the mean x. This would amount to saying, somewhat loosely, of a historical series, that it represented a phenomenon in a mature stage of development.

Let the position of the line of secular trend, Figure 1, be specified by two of the parameters:

a, the x-intercept; m, the slope; T, the negative t-intercept; which are related by the equation:

$$a = mT \tag{1}$$

Then, the equation of the line is:

$$x = m(i + T) \tag{2}$$

---

[1] Presented to the American Mathematical Society on December 28, 1921, under the title "Note on the determination of the rectilinear secular trend of an ordered series of statistical relatives."

[2] Warren M. Persons, *Review of Economic Statistics*, Jan., 1919, p. 12.

[3] *Ibid.*, p. 35, and April, 1919, p. 117 *et seq.*

the actual deviation of a particular variate, x_i, is:

$$x_i - m(i + T)$$

and the ratio of this deviation to x is:

$$\lambda_i = \frac{x_i}{m(i+T)} - 1. \tag{3}$$

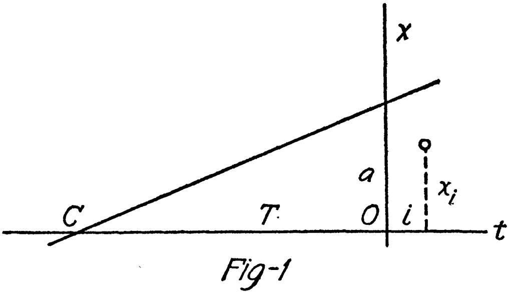

Fig-1

Because of the limitations placed upon the series, it is possible to expand in powers of $u_i = i/T$, and (3) becomes approximately:

$$\begin{aligned} \lambda_i &= \frac{x_i}{mT}(1 - u_i) - 1 \\ &= \frac{x_i}{a} - 1 - \frac{x_i}{a} u_i \\ &= P_i - p_i u_i \end{aligned} \tag{4}$$

where

$$p_i = x_i/a, \; P_i = p_i - 1.$$

Moreover

$$\lambda_i^2 = P_i^2 - 2P_i p_i u_i. \tag{5}$$

Suppose now that there is a second series of N variates, y_i, ordered relative to t_i, and subject to the same sort of limitations as the x series. We imagine that there is a certain "lag" between the two series, so that the t which belongs to the *ith* y differs from that belonging to the *ith* x by an amount L; and this lag, L, will be assumed constant. Then, if we take the y-axis at a distance L from the x-axis, values of x and y belonging to the *same* value of i will be associated with each other; and the analysis can proceed without further regard for the existence or amount of lag.

To distinguish between the two series, we suppose from now on that the above parameters for the x series are a_x, m_x, T_x, and that corresponding parameters for the y series are the same letters with subscript y. Moreover, for the y series, we use μ_i, v_i, q_i, and Q_i to correspond to λ_i, u_i, p_i, and P_i.

With this notation we get the following set of approximate relations:

$$\left.\begin{aligned}\lambda_i &= P_i - p_i u_i \quad &, \quad \mu_i &= Q_i - q_i v_i \\ \lambda_i^2 &= P_i^2 - 2\,P_i p_i u_i \quad &, \quad \mu_i^2 &= Q_i^2 - 2\,Q_i q_i v_i\end{aligned}\right\} \tag{6}$$

and the typical product term, which enters in the correlation coefficient, is:

$$\lambda_i \mu_i = P_i Q_i - Q_i p_i u_i - P_i q_i v_i + p_i q_i u_i v_i. \tag{7}$$

Whereas, by the assumptions about the series, p_i and q_i are not far from unity; u_i and v_i are of the first order in the small quantity—the ratio i/T—and it is possible that the same is true of P_i and Q_i, if a_x and a_y are not far from the arithmetic means. Thus all the terms in each of the expressions (6) and (7) are likely to be of the same order. If we were concerned with the actual deviations, the means, standard deviations, and correlation coefficient would depend upon the terms $P_i Q_i$, P_i^2, Q_i^2 and $P_i Q_i$. But, in the analysis from the point of view of the ratios λ and μ, the additional terms are likely to be equally important.

Granting that it is essential to keep all the terms in the above expressions, will their values be altered greatly if the lines of secular trend are determined by considering the ratios λ and μ, rather than the actual deviations? It is shown below that the values of a, as found by the two methods, do not differ greatly from each other. But the other parameter, T, and hence also m, is likely to be changed notably by applying the method of least squares to the ratios instead of the actual deviations. Hence, as T_x enters in u and T_y in v, the above values are all likely to suffer considerable change. It is therefore doubtful if it is warrantable to compute the correlation coefficient of the ratios unless the line of trend for each series is located on the basis of ratios.

II. FITTING THE LINE

We proceed next to fit a line to the x series by the method of least squares, basing the calculations upon the ratios instead of the actual deviations: we seek to determine the parameters a, m, and T so that the sum of the squares of the λ_i shall be minimum. The *necessary* conditions for a minimum are that the partial derivatives of $\Sigma\lambda_i^2$ with respect to m and T shall vanish.

The exact expression for λ_i is given in (3) as:

$$\lambda_i = \frac{x_i}{m(i+T)} - 1$$

from which

$$\frac{\partial\lambda_i}{\partial m} = -\frac{x_i}{i+T}, \quad \frac{\partial\lambda_i}{\partial T} = -\frac{x_i}{m}$$

giving

$$\frac{\partial \Sigma \lambda_i^2}{\partial m} = -2\sum \frac{x_i}{i+T}\lambda_i, \quad \frac{\partial \Sigma \lambda_i^2}{\partial T} = -2\sum \frac{x_i}{m}\lambda_i. \tag{8}$$

Each of these exact relations (8) is expanded in powers of u, and set equal to zero. Solving the two resulting equations simultaneously, and using (1), will give the following approximate values of the parameters:

$$m = \frac{1}{T}\,\frac{\Sigma x^2}{\Sigma x} \tag{9}$$

$$T = \frac{3\,\Sigma x \Sigma x^2 i^2 - 2\,\Sigma x^2 \Sigma x i^2 - \Sigma x i \Sigma x^2 i}{\Sigma x \Sigma x^2 i - \Sigma x^2 \Sigma x i} \tag{10}$$

$$a = \frac{\Sigma x^2}{\Sigma x} \tag{11}$$

The quantities involved in (9), (10), (11) are:

$$\Sigma x,\ \Sigma xi,\ \Sigma xi^2,\ \Sigma x^2,\ \Sigma x^2 i,\ \Sigma x^2 i^2$$

and these are the moments of orders 0, 1, 2, of the x and the x^2. The computation of the parameters which locate the line necessitates, then, the calculation of these six moments.

Example (i). Consider the following ordered series:

x	2	3	2	4	3	2	4
i	−3	−2	−1	0	1	2	3

The above formulas give for the intercepts:

$$a = 3.1; \quad T = 17.2$$

whereas computation by the usual method of least squares of the *deviations* gives:

$$a = 2.9; \quad T = 16.3.$$

The series is shown in Figure 2; and the two lines are plotted, the full line being that of the present method of calculation.

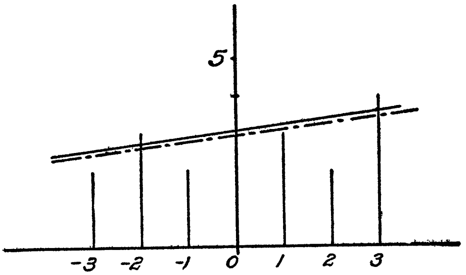

Fig-2

Example (ii). Examination of the ordered series:

x	9	9	10	8	9	11	10	12	11
i	–4	–3	–2	–1	0	1	2	3	4

gives:

$$a=10.0;\quad T=62.2$$

by the method of *ratios;* and, by the method of *deviations:*

$$a=9.9;\quad T=29.7.$$

This series and its two lines are shown in Figure 3.

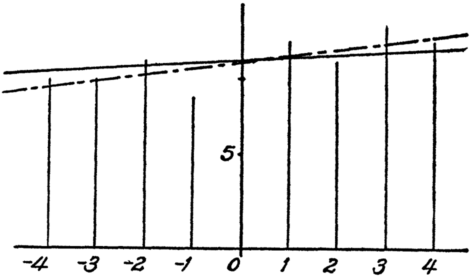

Fig-3

Example (iii). We study next a series of the historical economic type, the monthly values of building permits for twenty leading cities over a period of sixteen years as shown in the table[1] that follows. Calculation on the basis of ratios gives for the parameters:

$$a=440;\quad T=3130;\quad m=.14$$

whereas the determination depending on the actual deviations gives:

$$a=398;\quad T=1470;\quad m=.27$$

These last values are not the same as those found from the equation as given by Professor Persons,[2] but the discrepancy is due to the fact that he used only the first fourteen years of the series in fitting the line.

We observe that in each of the examples the values of a found by the two methods are not greatly different from each other, and that the value found by the ratio method exceeds that found by the older method. This is true in general, for:

$$a=\frac{\Sigma x^2}{\Sigma x}=a'\left\{1+\left(\frac{\sigma}{a'}\right)^2\right\}$$

where a' is the arithmetic mean, and σ is the standard deviation, of the x_i. But a' is precisely the value of a found by the older method;

---

[1] *Rev. of Econ. Stat.*, January, 1919, p. 74.

[2] *Ibid.*, p. 39.

	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18
January	169	169	234	346	357	160	359	334	345	276	314	298	267	349	343	212
February	201	133	289	317	332	168	428	337	272	313	335	313	301	320	324	173
March	416	317	522	494	510	279	616	609	510	497	467	547	464	549	495	190
April	328	396	495	617	673	422	714	614	543	650	548	489	499	563	476	286
May	343	333	461	530	632	355	662	518	427	594	506	506	531	673	412	259
June	310	399	546	543	555	464	621	501	564	576	490	519	402	636	368	243
July	282	388	461	491	456	422	607	433	499	483	444	506	405	880	340	246
August	266	313	485	478	387	338	489	426	631	520	393	342	448	393	290	225
September	189	355	463	439	336	365	487	414	466	368	391	339	435	400	292	185
October	246	312	377	458	378	473	418	419	407	411	366	311	438	583	235	118
November	207	309	381	397	231	446	381	428	379	385	295	218	419	425	248	81
December	182	258	292	339	217	341	372	365	336	468	438	237	443	442	170	

MONTHLY VALUES OF BUILDING PERMITS ISSUED FOR TWENTY LEADING CITIES

Original items in units of $100,000

and the ratio σ/a' is small because of the initial restrictions imposed upon the series.

On the other hand it is clear that the values of T may be closely alike or widely different, depending upon the form of the series. It is this fact which justifies the practical conclusion of Part I.

It has not been the purpose of this paper to secure greater refinement in the computation of the correlation coefficient for ordered series. Indeed, it is sought here to point out the lack of justification for as great refinement as is ordinarily used. The fact is that we are still far from our goal of securing a satisfactory index of the correspondence between two ordered series. It is worthy of note that in its magnificent work the Harvard Bureau uses the coefficient chiefly to determine the lag between two series. It has seemed advisable to work out the above relations; for, however uncertain is our knowledge of the exact portion of a particular variation in a historical series which is assignable to "secular growth," from a mathematical point of view the position of the line is one of the most precise things with which we deal. Further study of the line, as determined on this new principle, may increase our knowledge of the correlation between ordered series and point the way to a new method of separating out the "cyclical" portion of the variation in historical series, to examine which is usually our ultimate object.